Universal Solvation Model Based on Conductor-Like Screening Model

DEREK M. DOLNEY,[1] GREGORY D. HAWKINS,[1] PAUL WINGET,[1] DANIEL A. LIOTARD,[2] CHRISTOPHER J. CRAMER,[1] DONALD G. TRUHLAR[1]

[1]*Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431*

[2]*Laboratoire de Physico-Chimie Théorique, Université de Bordeaux 1, 351 Cours de la Liberation, 33405 Talence Cedex, France*

Received 6 July 1999; accepted 21 October 1999

ABSTRACT: Atomic surface tensions are parameterized for use with solvation models in which the electrostatic part of the calculation is based on the conductor-like screening model (COSMO) and the semiempirical molecular orbital methods AM1, PM3, and MNDO/d. The convergence of the calculated polarization free energies with respect to the numerical parameters of the electrostatic calculations is first examined. The accuracy and precision of the calculated values are improved significantly by adjusting two parameters that control the segmentation of the solvent-accessible surface that is used for the calculations. The accuracy of COSMO calculations is further improved by adopting an optimized set of empirical electrostatic atomic radii. Finally, the electrostatic calculation is combined with SM5-type atomic surface tension functionals that are used to compute the nonelectrostatic portions of the solvation free energy. All parameterizations are carried out using rigid (R) gas-phase geometries; this combination (SM5-type surface tensions, COSMO electrostatics, and rigid geometries) is called SM5CR. Six air–water and 76 water–solvent partition coefficients are added to the training set of air–solvent data points previously used to parameterize the SM5 suite of solvation models, thereby bringing the total number of data points in the training set to 2266. The model yields free energies of solvation and transfer with mean unsigned errors of 0.63, 0.59, and 0.61 kcal/mol for AM1, PM3, and MNDO/d, respectively, over all 2217 data points for neutral solutes in the training set and mean unsigned errors of 3.0, 2.7, and 3.1 kcal/mol, respectively, for 49 data points for the ions. J Comput Chem 21: 340–366, 2000

Correspondence to: D. G. Truhlar; e-mail: truhlar@umn.edu

Contract/grant sponsors: National Science Foundation; Alfred P. Sloan Foundation; Minnesota Supercomputer Institute; University of Minnesota

Keywords: solvation model; conductor-like screening model; atomic surface tensions; semiempirical molecular orbital methods

Introduction

In recent work our group developed five semiempirical solvation models for calculating free energies of solvation in water and virtually any organic solvent.[1–9] These are called universal solvation models (SM); in particular, SM5 universal solvation models denote that they are based on our fifth generation of functional forms[1, 2, 4, 6] for parameterizing atomic surface tensions.

The SM5.0R solvation model[3] is based entirely on atomic surface tensions, while the SM5.2R,[4] SM5.4,[1, 2, 5] SM5.42R,[6–8] and SM5.42[9] models combine atomic surface tensions with a quantum mechanical self-consistent reaction field (SCRF) electrostatic calculation[10–12] based on the generalized Born (GB) approximation.[13–16] The present article presents a new type of universal solvation model based on SM5-type functional forms for atomic surface tensions combined with a quantum mechanical SCRF electrostatic calculation based on the conductor-like screening model (COSMO) of Klamt and Schüürmann.[17] This combination is denoted SM5C (SM5 based on COSMO). The model is parameterized using rigid (gas-phase) solute geometries, which is indicated by appending an R (for rigid) to the model name, yielding SM5CR.

Three parameterizations of the SM5CR model are presented here: SM5CR/AM1, SM5CR/PM3, and SM5CR/MNDO/d. These parameterizations are intended for use with the semiempirical molecular orbital methods AM1,[18–20] PM3,[21] and MNDO/d,[22–24] respectively.

Theory

The SM5 solvation models are explained elsewhere,[1–9] and all aspects are the same here except for the way the electrostatic effects are calculated and some details of the functional forms and parameters as explained below. The most important change is that the COSMO algorithm is substituted for the GB algorithm in the electrostatic calculation. The theoretical formalism is reviewed here as briefly as possible, primarily to present the notation used in subsequent sections of the article.

SM5CR FORMALISM

The standard-state free energy of solvation (where the standard state is defined by the same concentration in the liquid solution and the vapor, e.g., 1 mol/L) is written as a sum of two terms[1, 2, 4, 5, 16, 25]:

$$\Delta G_S^0 = \Delta G_{\mathrm{EP}}\left(\{\rho_Z\}, \varepsilon\right) + G_{\mathrm{CDS}}\left(\left\{r_Z^{\mathrm{vdW}}\right\}, \Gamma\right), \quad (1)$$

where ΔG_{EP} is the electronic-polarization term, and G_{CDS} is the cavity-dispersion solvent-structure term. Note that ΔG_{EP} depends on a set of intrinsic electrostatic atomic radii $\{\rho_Z\}$ (where Z is an atomic number) and the solvent dielectric constant ε, and G_{CDS} depends on a set of van der Waals radii $\{\rho_Z^{\mathrm{vdW}}\}$ and a set of solvent descriptors $\Gamma = \{n, \alpha, \beta, \gamma, \phi, \psi, r^{\mathrm{SR}}, r^{\mathrm{MR}}\}$, which are further discussed below.

The first term in eq. (1) is called the electrostatic term, ΔG_{EP}, and it accounts for the distortion of the solute electronic (E) structure by the solvent and the electric polarization (P) of the solvent medium. In particular

$$\Delta G_{\mathrm{EP}}\left(\{\rho_Z\}, \varepsilon\right) = \Delta E_{\mathrm{E}}\left(\{\rho_Z\}, \varepsilon\right) + G_{\mathrm{P}}\left(\{\rho_Z\}, \varepsilon\right), \quad (2)$$

where ΔE_{E} is the electronic distortion energy of the solute and G_{P} is the electric-polarization free energy change due to insertion of the solute into the solvent, including the solvent reorganization cost. In general, ΔE_{E} is positive with solute distortion driven by the negative G_{P}. The G_{P} term is approximated here by the COSMO method described elsewhere.[17]

As in previous work,[1, 5] the term G_{CDS} is decomposed into short-range (SR) and medium-range (MR) effects:

$$G_{\mathrm{CDS}}\left(\left\{r_Z^{\mathrm{vdW}}\right\}, \Gamma\right) = \sum_k \sigma_k^{\mathrm{SR}}(\Gamma) A_k\left(\left\{r_Z^{\mathrm{vdW}}\right\}, r^{\mathrm{SR}}\right) + \sigma^{\mathrm{MR}}(\Gamma) \sum_k A_k\left(\left\{r_Z^{\mathrm{vdW}}\right\}, r^{\mathrm{MR}}\right), \quad (3)$$

where σ_k^{SR} is the short-range atomic surface tension of atom k, and A_k is its solvent-accessible surface area. The solvent-accessible surface area depends on the set of atomic van der Waals radii $\{r_Z^{\mathrm{vdW}}\}$, as well as two solvent radii, r^{SR} and r^{MR}. The radii $\{r_Z^{\mathrm{vdW}}\}$ were set equal to the van der Waals radii of Bondi.[26] The first sum of eq. (3), which represents the short-range effects, including dispersion,

depends on atom-specific contributions σ_k^{SR}. The second term, representing medium-range (MR) contributions, is a molecular contribution included in the case of nonaqueous solvents. The surface tensions depend on the solvent descriptors Γ, which is discussed below. Note that σ_k^{SR}, σ^{MR}, and the A_k also depend on the solute geometry (as does ΔG_{EP}), although this is not explicitly shown in the notation.

To compute σ_k^{SR}, the atomic surface tension term for atom k in eq. (3), SM5 solvation models utilize surface tension functional forms that depend on the geometry of the solute and Γ.[1–9] These functional forms employ a geometry-dependent switching function called a cutoff tanh (COT) function[2] and a set of underlying empirical parameters called surface tension coefficients. A COT function is expressed as

$$T\left(R_{kk'}\middle|\overline{R},\Delta R\right) = \begin{cases} \exp\left[-\left(\frac{\Delta R}{\Delta R - R_{kk'} + \overline{R}}\right)\right] & R_{kk'} \le \overline{R} + \Delta R, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where $R_{kk'}$ is the distance between two atoms k and k', and $\overline{R}$ is the midpoint of the switch. Note that the range over which the function switches is $2\Delta R$.

The present model adopts the improved SM5 functional forms,[4,6] which are given by the following equations:

$$\sigma_k^{SR}\big|_{k=H} = \tilde{\sigma}_H + \sum_{k'=C,N,O,S}\left\{T\left(R_{kk'}\middle|\overline{R}_{Hk'},W\right) \times\left[\tilde{\sigma}_{Hk'} + \tilde{\sigma}_{HN}^{(2)}\sum_{\substack{k''=N\\k'=N\\k'\neq k''}} T\left(R_{k'k''}\middle|\overline{R}_{NN},W\right) + \tilde{\sigma}_{HO}^{(2)}\sum_{\substack{k''=H\\k''\neq k\\k'=O}} T\left(R_{k'k''}\middle|\overline{R}_{OH},W\right)\right]\right\}, \quad (5)$$

$$\sigma_k^{SR}\big|_{k=C} = \tilde{\sigma}_C + \tilde{\sigma}_{CC}\sum_{\substack{k'=C\\k'\neq k}} T\left(R_{kk'}\middle|\overline{R}_{CC},W\right) + \tilde{\sigma}_{CC}^{(2)}\sum_{\substack{k'=C\\k'\neq k}} T\left(R_{kk'}\middle|\overline{R}_{CC}^{(2)},W_{CC}\right) + \tilde{\sigma}_{CN}\left[\sum_{k'=N} T\left(R_{kk'}\middle|\overline{R}_{CN},W\right)\right]^x, \quad (6)$$

$$\sigma_k^{SR}\big|_{k=N} = \tilde{\sigma}_N + \tilde{\sigma}_{NC}\left\{\sum_{k'=C} T\left(R_{kk'}\middle|\overline{R}_{CN},W\right) \times\left[\sum_{\substack{k''\neq k\\k''\neq k'}} T\left(R_{k'k''}\middle|\overline{R}_{Ck''},W\right)\right]^y\right\}^z + \tilde{\sigma}_{NC}^{(2)}\sum_{k'=C}\left[T\left(R_{kk'}\middle|\overline{R}_{CN},W\right) \times\sum_{k''=O} T\left(R_{k'k''}\middle|\overline{R}_{CO},W\right)\right] + \tilde{\sigma}_{NC}^{(3)}\sum_{k'=C} T\left(R_{kk'}\middle|\overline{R}_{OC}^{(2)},W_{NC}\right), \quad (7)$$

$$\sigma_k^{SR}\big|_{k=O} = \tilde{\sigma}_O + \tilde{\sigma}_{OC}\sum_{k'=C} T\left(R_{kk'}\middle|\overline{R}_{OC}^{(2)},W_{OC}\right) + \tilde{\sigma}_{ON}\sum_{k'=N} T\left(R_{kk'}\middle|\overline{R}_{ON},W\right) + \tilde{\sigma}_{OO}T\left(-\sum_{\substack{k'=O\\k'\neq k}} T\left(R_{kk'}\middle|\overline{R}_{OO},W\right)\middle|\overline{R}_{TT},W_{TT}\right) + \tilde{\sigma}_{OP}\sum_{k'=P} T\left(R_{kk'}\middle|\overline{R}_{OP},W\right), \quad (8)$$

$$\sigma_k^{SR}\big|_{k=F} = \tilde{\sigma}_F, \quad (9)$$

$$\sigma_k^{SR}\big|_{k=P} = \tilde{\sigma}_P, \quad (10)$$

$$\sigma_k^{SR}\big|_{k=S} = \tilde{\sigma}_S + \tilde{\sigma}_{SS}\sum_{\substack{k'=S\\k'\neq k}} T\left(R_{kk'}\middle|\overline{R}_{SS},W\right) + \tilde{\sigma}_{SP}\sum_{k'=P} T\left(R_{kk'}\middle|\overline{R}_{SP},W\right), \quad (11)$$

$$\sigma_k^{SR}\big|_{k=Cl} = \tilde{\sigma}_{Cl}, \quad (12)$$

$$\sigma_k^{SR}\big|_{k=Br} = \tilde{\sigma}_{Br}, \quad (13)$$

and

$$\sigma_k^{SR}\big|_{k=I} = \tilde{\sigma}_I, \quad (14)$$

where $\tilde{\sigma}_k$ and $\tilde{\sigma}_{kk'}$ are the empirical surface tension coefficients for atom k or atom pair kk'. The three exponents in eqs. (6) and (7), labeled x, y, and z, were given the values 2, 2, and 1.3, respectively, in all previous SM5 models that employ these functional forms.[4,6] The reoptimization of these parameters for the SM5CR models is discussed in the parameterization section.

The short-range surface tension coefficients $\tilde{\sigma}_k$ and $\tilde{\sigma}_{kk'}$ are constant for solutes in water, but they depend on solvent properties in the case of nonaqueous solvents. The general form is a modified version of that[4,5] used previously; in particular, the form for SM5CR is expressed as

$$\tilde{\sigma}_k = \hat{\sigma}_k^{(n)}\left(1 - \frac{1}{n^2}\right) + \hat{\sigma}_k^{(\alpha)}\alpha + \hat{\sigma}_k^{(\beta)}\beta, \quad (15)$$

where n is again the index of refraction n_D^{20} for sodium D line radiation at 20°C,[27] and α and β are Abraham's hydrogen-bond acidity and basicity

descriptors, respectively.[28–30] (In Abraham's notation α is $\sum \alpha_2^H$ and β is $\sum \beta_2^H$.) The coefficients $\hat{\sigma}_k$ and $\hat{\sigma}_{kk'}$ are obtained by least squares minimization in the parameterization of SM5CR for nonaqueous solvents. The modification from previous work[1–9] is that in the SM5CR model the n is replaced by $(n^2 - 1)/n^2$, which provides a better measure of the electronic contribution to the electric susceptibility of the solvent and hence a better measure of the strength of the dispersion forces. This change reduced the mean unsigned error over the 2217 solvation and transfer free energies of neutral solutes in the training set (described in a later section) by a mere 0.01 kcal/mol, but more importantly, the functional form $(n^2 - 1)/n^2$ represents the optical-frequency version of the Born function $(\varepsilon - 1)/\varepsilon$ and should thus impart more physical behavior to the SM5CR model.

Finally, the medium-range surface tension, σ^{MR} in eq. (3), is computed as[4]

$$\sigma^{MR} = \hat{\sigma}_{MR}^{(\gamma)}\gamma + \hat{\sigma}_{MR}^{(\beta^2)}\beta^2 + \hat{\sigma}_{MR}^{(\phi^2)}\phi^2 + \hat{\sigma}_{MR}^{(\psi^2)}\psi^2, \quad (16)$$

which depends on the macroscopic surface tension (γ),[27] the square of the hydrogen-bond basicity descriptor (β^2), the square of the fraction of nonhydrogenic solvent atoms that are aromatic carbon atoms (ϕ^2), and the square of the fraction of nonhydrogenic solvent atoms that are electronegative halogen atoms (ψ^2). The medium-range surface tension is omitted for the solvent water; the physical effects it represents are included in the first sum of eq. (3) for water as solvent, so σ_k^{SR} absorbs the medium-range effects as well, and r^{MR} does not appear in the calculations for water.

The values of the solvent descriptors used to parameterize the SM5CR model are presented in Table S-I. These are the same values as those that were used to parameterize the SM5.0R, SM5.2R, SM5.4, and SM5.42R models. (The SM5.42 model uses the same parameters as SM5.42R.)

COMPUTATIONAL METHODS: ELECTROSTATIC CONTRIBUTION

To calculate G_P, the COSMO method defines a solute surface (SS) as the envelope of a set of overlapping nuclear-centered spheres from which small regions in the vicinity of sphere–sphere intersections have been removed. The solvent is first approximated as a conductor of infinite extent, which allows the electric polarization of the solvent dielectric medium in the presence of the solute to be represented by a set of effective partial charges on the SS. After calculating ΔG_{EP} for $\varepsilon = \infty$, the result is scaled (in the original implementation[17]) to approximate the result for finite dielectric constants ε by the Onsager factor $f_O(\varepsilon) = (\varepsilon - 1)/(\varepsilon + 0.5)$. The theory is documented elsewhere.[17] Past SM5 models, however, used the GB approximation to calculate G_P, which employs a similar scaling factor $f_B(\varepsilon) = (\varepsilon - 1)/\varepsilon$. The Born form of $f(\varepsilon)$ is appropriate for monopoles, and the factor $f_O(\varepsilon)$ applies to pure dipoles. We decided to continue using the form $f_B(\varepsilon)$ in the SM5CR formalism. By substituting $f_B(\varepsilon)$ for $f_O(\varepsilon)$ in COSMO, the overall mean unsigned error over the 2217 solvation and transfer free energies of neutral solutes in the training set is reduced by 0.02 kcal/mol.

The calculation of ΔG_{EP} involves effective partial charges that are placed on segments of the SS. Details regarding segmentation of the SS into locations of discrete point charges cannot be found in the literature, but in practice the standard COSMO method is defined by the algorithms in the two available computer programs in which semiempirical COSMO is available.[31, 32] The present work revealed that the grids used by default in those programs are inadequate for many applications. Thus, the standard values of some keyword-controlled parameters in these programs were modified, which will now be discussed.

The COSMO method uses a tessellation scheme similar to GEPOL[33] to define a segmentation on the SS. In order to balance accuracy and computational efficiency, the COSMO method calculates electrostatic interaction energies between pairs of segments using two distinct distributions of points on the SS. These distributions differ in the density of points, and they will be referred to as the coarse grid and the fine grid. In the COSMO method a segment is defined as a single coarse grid point along with a set of fine grid points associated with it in a manner to be explained shortly. At sufficient distance the interaction between pairs of segments is calculated simply as the Coulomb energy of point charges whose location coincides with the coarse grid points of the segments. At distances below this cutoff value, the interaction between two segments is computed by distributing the charge on each segment uniformly over its fine grid points and summing over pairwise Coulomb interactions of the fine grid point charges.

Clearly, accuracy and computation time are dependent on the point densities of the coarse and fine grids, as well as the cutoff value that determines which grid will be used to calculate each pairwise segment interaction. In the available implementations[31, 32] the fine grid is constructed by

starting with 1082 points per atom-centered sphere and eliminating those that lie within spheres of other atoms. Note that the initial 1082 points are the vertices of a regular polyhedron that is used to "uniformly" sample the sphere of the atom under construction. The number of coarse grid points is also restricted to be the number of vertices of a regular polyhedron and is therefore chosen from among the values 12, 32, 42, 92, 122, 162, 272, 362, and so forth. The actual number of coarse grid points that are used depends on the value of an integer keyword called NSPA, which defaults to 60 in MOPAC93,[31] 30 in early versions of Semichem's AMPAC,[32] and 42 in Semichem's AMPAC after July 1998 according to Klamt's recommendations of a value between 42 and 92. Intermediate values of NSPA (values among 12, 32, 42, 92, etc.) are used in the code for a best "smoothing" of the number of segments with changes in geometry; some vertices of the polyhedrons become exposed or not. The initial number of coarse grid points placed on the sphere depends not only on NSPA but on the atom type; a smaller number of coarse grid points may be used for hydrogen atoms than for heavier atoms. The early AMPAC default of 30 for NSPA results initially in 12 coarse grid points, regardless of atom type, and the latter AMPAC default and the MOPAC93 default of 60 both initially give 12 points on hydrogen-centered spheres and 42 points on all heavier atoms. In the present work we concluded that the best compromise between precision and computational cost is achieved by setting NSPA to 90. This value initially yields 12 coarse grid points per hydrogen atom and 42 per heavier atom.

Whichever value is used for NSPA, the initial number of coarse grid points is refined by the NSPA-dependent iterative procedure described below. Once the segmentation is converged, the number of coarse grid points may differ from the initial value.

There is another integer keyword called DISEX, which is a distance. When two segments are sufficiently separated from each other, their mutual interaction may be approximated by a two-point interaction because each segment, although actually a small area, becomes effectively equivalent to its geometrical center with all its charge being concentrated at that one point. At short distances (shorter than DISEX) no such averaging is carried out; instead the exact sum of interactions is carried out using all the points on the fine grids of both segments. Thus, a change in the value of DISEX will change the accuracy with which the matrix elements are evaluated but will not change the number of segments or the size of the matrix.

Although COSMO gradients are not computed in the present article, it is useful for the comparison of methods in the Discussion to note that the "analytical" calculation of the gradient in COSMO is based entirely on the course grid averaging approximation (the two-point charges approximation). With large values of DISEX there are more matrix elements calculated without such averaging, and so the gradient obtained this way will show larger inconsistencies with the energy calculation.

The distribution of fine grid points and coarse grid points resulting from the choices of NSPA defines the initial state for the following iterative method:

1. Each fine grid point on the SS is associated with a coarse grid point that lies on the same atom-centered sphere. The program defines an angular separation threshold that depends on NSPA. If a fine grid point does not lie within the threshold of a coarse grid point, it is added to the coarse grid. Hence, for some values of NSPA, the final grid may be larger than the initial coarse grid.
2. Once the association is complete, any coarse grid points that were not associated with at least one fine grid point are eliminated from the coarse grid. At this point, each coarse grid point has an associated (nonempty) set of fine grid points. This defines a segmentation of the SS.
3. For each segment, the position vectors representing the location of its fine grid points are averaged. These vector averages become the new coarse grid point distribution.

The three steps are repeated until a convergence condition is met (i.e., a negligible change in coarse grid position vectors between successive iterations). At that stage a converged coarse grid is defined on the SS, and each fine grid point is associated with exactly one point of the coarse grid. This defines the segmentation used to calculate the polarization energy of the solvent.

It was determined in the present work that better results can be obtained by increasing the value of DISEX to 4, which results in a larger number of segment interactions calculated using the fine grid and hence a more accurate numerical solution for G_P.

The increase in computation time resulting from greater segment density was justified by the increase in the accuracy and precision of calculated

solvation free energies, particularly in the case of ions, where polarization effects dominate the solvation free energy. Table I shows the values of ΔG_{EP} calculated for several molecules and ions in water with four selections of NSPA and DISEX parameters. The converged value was obtained by increasing NSPA to very large values such that the total number of segments on the SS was about 1000, as shown in Table I. At very large numbers of segments the ΔG_{EP} was found to converge to a stable value. In principle, the computation time would vary approximately as NSPA cubed because a matrix (the representation of the Coulomb operator in the basis of the segments) whose size is of the order of NSPA must be inverted; in practice, however, the scaling is much less severe due to overhead, although computation time is excessive for NSPA on the order of 1000. For the four more practical choices of NSPA, Table I displays the error in the computed ΔG_{EP} with respect to this converged value.

The results in Table I can be summarized as follows. The grid parameters in AMPAC yield a mean unsigned relative error of 4.6% at a mean computer time of 0.30 s. The grid parameters in MOPAC decrease the mean unsigned relative error to 2.0% at a mean computer time of 0.33 s. The grid parameters recommended above yield a mean unsigned relative error of only 0.9% in the electrostatic calculation by increasing the mean computer time to 0.41 s. The new tighter grids were used for parameterization to reduce the computational noise.

Parameterization

The procedure used to parameterize the SM5CR models is based on the parameterization procedure used for the SM5.2R model.[4] The most important deviation from past parameterizations is that water–organic solvent partition coefficient data points have been added to the training set. The training set is broadened significantly for solutes containing N and P. Other changes to the procedure are minor, but they are presented below for completeness.

PARAMETERIZATION PROCEDURE

The SM5CR model has four types of parameters: the linear surface tension coefficient parameters [i.e., the $\hat{\sigma}_k$, $\hat{\sigma}_{kk'}$, and the $\hat{\sigma}_{MR}^{(\Gamma)}$ parameters in eqs. (15) and (16)]; the electrostatic radii, $\{\rho_Z\}$ in eq. (2), for each atomic number; the effective solvent radii, r^{SR} and r^{MR} of eq. (3); and nonlinear parameters in the surface tension functionals, eqs. (5)–(14), which include the exponents x, y, and z, in addition to the COT parameters. This order roughly reflects the sensitivity of the model to the values selected for the parameters in each category.

The linear parameters (the first type) were optimized using linear regression and were always refit after any of the other parameters were adjusted. The procedure is unchanged from previous parameterizations[4] in that the linear parameters are fit in three stages: solutes containing only H, C, N, and O atoms are used to optimize the surface tension coefficients corresponding to those atoms; solutes containing sulfur and halogen atoms are used to fit their corresponding surface tension coefficients; and solutes containing phosphorus are used to fit phosphorus surface tension coefficients. However, because of the addition of water–organic solvent partition coefficients, the air–water and air–organic solvent data can no longer be fit sequentially, and all 2217 pieces of data for neutral solutes are fit simultaneously.

It was discovered that the accuracy of ΔG_{EP} calculations by COSMO could be significantly improved by modifying the electrostatic radii used by COSMO to define the SS. Table II shows several sets of electrostatic radii and the value of the unfitness function, U, developed in earlier work,[34]

$$U = \frac{1}{N+I}\left(\sum_{n=1}^{N}\left|G^0_{S,\mathrm{exp},n} - G^0_{S,\mathrm{calc},n}\right| + \frac{1}{6}\sum_{i=1}^{I}\left|G^0_{S,\mathrm{exp},i} - G^0_{S,\mathrm{calc},i}\right|\right), \quad (17)$$

where N is the total number of neutral molecules and I is the total number of ionic compounds in the training subset [i.e., the subsets defined in the preceding paragraph]. The $G^0_{S,\mathrm{exp}}$ is the experimental standard-state free energy of solvation, and $G^0_{S,\mathrm{calc}}$ is the standard-state free energy of solvation calculated with the selected set of nonlinear parameters and reoptimized set of nonlinear parameters.

The electrostatic radii for the SM5CR model were optimized by separately examining the performance of SM5CR/AM1 over the aqueous solvation free energy of the subsets of the training set. The values chosen for the SM5CR model are presented in Table II. The radii for H, N, and O were adjusted to minimize the error in the calculated solvation free energy for the five ions displayed in the table. These ions will improve the precision of the model in predicting solvation free energies for other compounds with similar charged centers (e.g., a model that calculates the solvation free energy of NH_4^+ accurately could also be expected to yield reasonably

TABLE I.

Computed ΔG_{EP} in Water with COSMO at Various Settings of NSPA and DISEX Using MNDO/d Hamiltonian and SM5.2R Electrostatic Radii.

Solute	NSPA	DISEX	Total SS Segments	ΔG_{EP} (kcal/mol)	Error Absolute	Error %	Computation Time[a] (s)
F^-	30	2.0	12	−110.46	.70	1.6	0.20
	60	2.0	32	−110.25	−1.50	1.4	0.23
	90	4.0	43	−109.00	−0.24	0.2	0.27
	120	4.0	54	−109.10	−0.35	0.3	0.26
	1082	36.0	1082	−108.76			24.53
OH^-	30	2.0	17	−105.20	−2.14	2.1	0.22
	60	2.0	47	−104.63	−1.57	1.5	0.20
	90	4.0	56	−103.51	−0.45	0.4	0.26
	120	4.0	83	−103.53	−0.47	0.5	0.29
	1082	34.0	1024	−103.06			23.52
Methoxide	30	2.0	40	−85.06	−0.41	0.5	0.26
	60	2.0	90	−85.07	−0.42	0.5	0.22
	90	4.0	104	−84.49	0.16	−0.2	0.31
	120	4.0	142	−84.59	0.06	−0.1	0.31
	1082	30.0	1025	−84.65			24.22
Methylammonium	30	2.0	46	−78.66	−0.97	1.2	0.26
	60	2.0	83	−78.36	−0.66	0.9	0.24
	90	4.0	95	−77.93	−0.24	0.3	0.28
	120	4.0	134	−77.93	−0.23	0.3	0.30
	1082	36.0	1063	−77.69			25.12
Acetate	30	2.0	56	−76.38	−0.51	0.7	0.29
	60	2.0	97	−76.34	−0.46	0.6	0.23
	90	4.0	119	−75.86	0.02	0.0	0.34
	120	4.0	154	−75.95	−0.08	0.1	0.33
	1082	33.0	1067	−75.87			28.42
$AcetamideH^+$	30	2.0	60	−78.03	−1.59	2.1	0.29
	60	2.0	103	−77.10	−0.65	0.9	0.30
	90	4.0	124	−76.51	−0.06	0.1	0.34
	120	4.0	165	−76.54	−0.09	0.1	0.37
	1000	34.0	898	−76.45			16.91
Glycine zwitterion	30	2.0	59	−52.89	−1.42	2.8	0.29
	60	2.0	103	−52.92	−1.44	2.8	0.29
	90	4.0	128	−51.97	−0.50	1.0	0.38
	120	4.0	170	−51.95	−0.48	0.9	0.42
	1082	41.0	1080	−51.47			29.63
n-Propylamine	30	2.0	81	−4.44	−0.70	18.8	0.30
	60	2.0	147	−3.94	−0.21	5.5	0.31
	90	4.0	164	−3.87	−0.13	3.5	0.40
	120	4.0	219	−3.81	−0.08	2.1	0.50
	700	34.0	1081	−3.74			29.36
Piperidine	30	2.0	111	−3.43	−0.30	9.6	0.37
	60	2.0	194	−3.22	−0.09	2.8	0.46
	90	4.0	213	−3.18	−0.05	1.6	0.56
	120	4.0	286	−3.16	−0.03	0.8	0.86
	570	27.0	1082	−3.13			32.75

TABLE I.
(Continued)

Solute	NSPA	DISEX	Total SS Segments	ΔG_{EP} (kcal/mol)	Error Absolute	Error %	Computation Time[a] (s)
9-Methyladenine	30	2.0	111	−13.60	−0.97	7.7	0.44
	60	2.0	177	−13.13	−0.50	3.9	0.51
	90	4.0	206	−12.92	−0.29	2.3	0.66
	120	4.0	254	−12.86	−0.23	1.8	0.88
	680	43.0	1070	−12.63			34.16
Thiophenol	30	2.0	93	−5.10	0.39	−7.1	0.33
	60	2.0	152	−5.34	0.14	−2.6	0.37
	90	4.0	188	−5.39	0.09	−1.7	0.44
	120	4.0	235	−5.34	0.15	−2.7	0.60
	710	39.0	1077	−5.48			33.22
Dimethyldisulfide	30	2.0	70	−5.61	−0.27	5.0	0.30
	60	2.0	136	−5.43	−0.09	1.8	0.28
	90	4.0	158	−5.34	0.00	−0.1	0.38
	120	4.0	202	−5.38	−0.04	0.7	0.41
	810	38.0	1060	−5.34			27.30
1-Iodobutane	30	2.0	96	−1.76	−0.02	1.2	0.36
	60	2.0	169	−1.76	−0.02	1.2	0.34
	90	4.0	188	−1.73	0.01	−0.8	0.48
	120	4.0	249	−1.73	0.02	−1.0	0.62
	640	43.0	998	−1.74			26.59

The errors are relative to the converged value.
[a] Total CPU seconds on an IBM SP with 334-MHz 604e processors.

accurate values for protonated amines in general). The performance of the model was not sensitive to the carbon radius, and so the value used in the SM5.42R/AM1 model was used without alteration. The radius for *S* was changed. The SM5.42R radii for F, Cl, Br, and I give the smallest *U* over the halogen containing subset as compared to the default COSMO electrostatic radii and Bondi[26] van der Waals radii. Finally, the value of 2.56 Å for the phosphorus radius was found to give the smallest *U* over the third subset. Note that the calculations in Table II were performed after the third and fourth type parameters were selected, which is discussed next.

The effective solvent radii have been modified several times in the evolution of the SM*x* solvation models. We found in the development of SM5.2R[4] that effective solvent radii of zero provide a better fit to the available experimental data. However, subsequent unpublished work[35] confirmed that nonzero radii used earlier give results in better agreement with explicit solvent studies for the potential of mean force between two solutes. In the present work we found that zero solvent radii give an unfitness $U = 0.54$ kcal/mol, while setting $r^{SR} = 1.7$ Å and $r^{MR} = 3.4$ Å gives $U = 0.64$ kcal/mol. Although the overall errors are larger, the nonzero radii were chosen for the SM5CR model because they are thought to better describe the true physical nature of the problem. The use of two distinct effective solvent radii to distinguish short- and long-range solvation effects was justified in the developments of other SM*x* models.[5] In the present case, taking $r^{SR} = r^{MR} = 1.7$ Å results in $U = 0.68$ kcal/mol, confirming that the use of two solvent radii better correlates the experimental data. Recall that the σ^{MR} term in eq. (3) is always taken to be zero for aqueous solutes, which is also justified in previous articles.[5]

With regard to the fourth type of parameters (i.e., the nonlinear parameters in the surface tension functionals), the performance of the SM5CR model was evaluated with respect to the values of x, y, and z [cf. eqs. (6), (7)]. The value of z was changed from 1.3 to 2, resulting in a slight decrease in the overall U of 0.03 kcal/mol. No significant benefit could be realized by adjusting x and y, and so these values were not changed from the values used in other SM5

TABLE II.
Sets of Electrostatic Radii ρ_Z (Å) with Corresponding Values of Unfitness Function, *U*, and Errors in Calculated Solvation Free Energies (kcal/mol) for Ions into Water (All Calculated with SM5CR/AM1).

	Z	COSMO Defaults[31, 32]	Bondi Radii[26]	SM5.2R/AM1 Radii[4]	SM5CR Radii
	H	1.08	1.20	0.91	0.87
	C	1.53	1.70	1.78	1.78
	N	1.48	1.55	1.92	1.98
	O	1.36	1.52	1.60	1.47
	F	1.30	1.47	1.50	1.50
	P	1.75	1.80	2.40	2.56
	S	1.70	1.80	2.05	2.15
	Cl	1.65	1.75	2.13	2.13
	Br	1.80	1.85	2.31	2.31
	I	2.05	1.98	2.66	2.66
U (kcal/mol)[a]	(1) C, H, O, N	0.50	0.51	0.52	0.47
	(2) S, F, Cl, Br, I	0.96	0.71	0.52	0.51
	(3) P	2.38	1.71	1.09	1.22
$G^0_{S,\mathrm{exp}} - G^0_{S,\mathrm{calc}}$ (kcal/mol)	H_3O^+	13.2	20.0	7.9	3.7
	OH^-	−11.8	−2.2	3.3	−4.0
	NH_4^+	−6.6	−3.1	−5.8	−4.0
	$CH_3OH_2^+$	6.8	11.9	2.8	0.3
	$CH_3NH_3^+$	−3.1	0.1	−3.9	−2.6

The last column gives the radii selected for the SM5CR models.
[a] The training subset used to calculate this *U* includes all aqueous G^0_S data for neutral solutes and 43 ions. These 43 ions include all ions presented in Table VIII except $C_6H_5NH_3^+$, imidazoleH^+, $HC(OH)NH_2^+$, and CH_3CNH^+.

models. Likewise, the COT parameters were taken from their original parameterization in the development of the SM5.4[2] and SM5.2R[4] models.

ADDITIONS TO TRAINING SET

During the development of the SM5CR model we decided to augment the standard training set used to parameterize SM*x* models with additional data points. SM5 models previously treated water as a separate solvent, resulting in different sets of surface tension coefficients: one set for water and one set for organic solvents. In the past the training set included only absolute solvation free energies, so that the regression fits of water and of organic solvent data were statistically uncorrelated. However, there are a number of interesting solutes for which solvent–solvent partition coefficients are available but absolute free energies of solvation are not. In an effort to increase the robustness of the model, 76 solvent–solvent partition coefficient data were added to the training set, hence correlating the water and organic surface tension coefficients. Six more absolute free energies of solvation were also added. Thus, the previous training set contained 2135 absolute free energies of solvation for 275 neutral solutes in 91 solvents and no liquid–liquid partitioning data. The present training set contains 2141 absolute free energies of solvation for 278 neutral solutes in 91 solvents and 76 partitioning data for 54 more neutral solutes corresponding to transfer free energies between water and 12 of the original 90 organic solvents.

The additions to the training set were taken from two sources: the Medchem data base[36] and an article by Bona et. al.[37] Data from the Medchem data base were prescreened and only used if none of the following impediments was involved: the data point was measured outside the temperature range of 20–30°C, the data point was measured outside the pH range of 6–8, the data point was measured in an aqueous phase that was not purely aqueous, the solute was not in its true form, salting out occurred during the measurement, or the data point was marked as unreliable. For compounds that had more than one piece of data in a given solvent, an average value was calculated while removing data

that were more than 2 standard deviations from the mean iteratively until no more outliers existed.

It is possible to have several conformational minima for a given molecule. In general, these conformations need to be statistically averaged to obtain the true free energy of solvation. A simplification is the case that one conformation dominates the equilibrium in the gas phase and in solution. The dominant conformation of each molecule was determined as described previously.[2,38] Specific examples of this are conformations minimizing steric interactions were chosen for amides and ureas; the relative conformations of substituents in 1,3- or 1,4-disubstituted aromatics, substituents on separate rings in a bicyclic structure, or methyl rotamers were taken to be those found in the lowest energy conformation from AM1 calculations. Possible tautomeric equilibria were also considered, and only compounds for which previous calculations or experiments indicated a single dominant tautomer were used.[39–41]

The final list of solutes contains molecules composed of H, C, N, O, F, S, P, Cl, Br, and I. The new free energies of solvation are for 3,4-dimethylpyridine, 3,5-dimethylpyridine, and 4-ethylpyridine in water and in *n*-hexadecane. The 76 new log *P* data points are given in Table S-II of the supporting information. Note that the total number of ions is 49 and that the ion data are not used to fit surface tension coefficients; ion data are only used to calculate the unfitness function [eq. (17)].

HF/MIDI! solute geometries were invariably used for all solutes during the parameterization and for all SM5CR results in this article.

Results

Table III presents the surface tension coefficients optimized for the SM5CR/AM1, SM5CR/PM3, and SM5CR/MNDO/d models. The mean signed and unsigned errors for the three models are organized by solute class in Table IV and by solvent in Table V. Note that Table IV includes solutes containing phosphorus, but Table V excludes all solutes that contain phosphorus. Additionally, the new log *P* data points are distributed among the solute classes given in Table IV, but all log *P* data points occur on a separate line in Table V (i.e., they are not distributed among the solvent classes).

Table VI gives the solvation free energies calculated by the three methods for several specific solutes in water and various organic solvents. In addition, the ΔG_{EP} and G_{CDS} contributions to the solvation free energy for selected solutes in water and chloroform are presented in Table VII. This table makes apparent the fact that, in the case of neutral solutes, the electrostatic and nonelectrostatic contributions must both be included to properly predict the general trends in the solvation free energy. The solvation free energies of ions are presented in Table VIII.

For all neutral solutes, the SM5CR/AM1 model affords a mean unsigned error of 0.59 kcal/mol over the 2157 neutral data points that do not contain phosphorus and 0.63 kcal/mol when phosphorus data are included. The SM5CR/PM3 model gives overall mean unsigned errors of 0.56 and 0.59 kcal/mol with and without phosphorus data, respectively, and the SM5CR/MNDO/d model gives 0.59 and 0.61 kcal/mol for the respective mean errors.

Discussion

Comparing the performance of the three different semiempirical molecular orbital methods used here, one finds similar overall performance. In certain cases, one of the three methods performs significantly better for a given solute or solvent class, as may be determined by examining Tables IV and V. For example, PM3 corrects a systematic oversolvation of ethers present in the other two models. Encouragingly, MNDO/d predicts more accurate solvation free energies of phosphorus compounds, presumably owing to the inclusion of d orbitals into the MNDO/d Hamiltonian. The largest systematic errors for specific solvent classes occur for ketones, aromatic ethers, nitriles, nitrohydrocarbons, and haloaromatics. Table VI shows that the three parameterizations tend to follow the same general trends for specific solute–solvent combinations.

Note that MNDO/d reduces exactly to MNDO as long as no element heavier than F is involved.[24] So for molecules composed of such atoms, our parameterization is valid for the original MNDO method, as well as for MNDO/d.

The overall error is larger than for other SM5 models,[1–9] which is due in part to the larger training set and also to the reintroduction of the nonzero solvent radius as discussed previously. Although the overall errors are larger, it is believed that the behavior of the model is more physical because the additional training set data were used and the choice of solvent radii and index of refraction dependence is more physical.

In the original COSMO article[17] Klamt and Schüürmann argue that the relative error in the

TABLE III.

Surface Tension Coefficients (cal mol^{-1} Å^{-2}) for SM5CR Models.

	SM5CR/AM1				SM5CR/PM3				SM5CR/MNDO/d			
k	$\hat{\sigma}_k^{(n)}$	$\hat{\sigma}_k^{(\alpha)}$	$\hat{\sigma}_k^{(\beta)}$	$\hat{\sigma}_k^{(\text{water})}$	$\hat{\sigma}_k^{(n)}$	$\hat{\sigma}_k^{(\alpha)}$	$\hat{\sigma}_k^{(\beta)}$	$\hat{\sigma}_k^{(\text{water})}$	$\hat{\sigma}_k^{(n)}$	$\hat{\sigma}_k^{(\alpha)}$	$\hat{\sigma}_k^{(\beta)}$	$\hat{\sigma}_k^{(\text{water})}$
H	−4.60			19.73	−2.12			19.71	−1.61			19.69
C	99.89	22.18	10.15	46.10	105.75	21.63	0.20	45.55	119.70	16.39	−8.45	47.14
N	−36.34	−42.94	20.98	−9.88	−53.59	−52.59	37.08	−25.89	−18.32	−53.58	33.39	−2.56
O	65.51	19.85		11.49	22.54	8.43		−23.92	58.16	18.35		1.84
F	−2.74			21.46	−0.26			21.07	16.05			30.34
P	92.41			27.05	24.55			−12.88	19.52			−16.82
S	−63.81	4.50	35.49	11.77	−52.28	9.47	33.99	23.36	−79.89	−16.52	25.27	−12.60
Cl	−43.79			4.12	−45.44			1.36	−42.30			3.59
Br	−52.83			2.30	−52.66			1.54	−52.08			2.47
I	−57.65			−1.57	−49.71			4.45	−51.95			1.12
kk′												
H, C	−52.52			−19.76	−55.61			−21.34	−57.80			−22.59
H, N	11.56		−18.65	−27.79	−8.52		−36.26	−43.15	−9.83		−30.73	−43.15
H, N (2)	−139.78			−71.98	−119.79			−58.04	−141.93			−72.74
H, O	−45.47	−80.72	−24.22	−63.52	3.09	−75.51	−32.91	−25.02	−43.98	−89.50	−38.10	−63.09
H, O (2)	98.44			113.97	83.86			98.78	86.72			99.63
H, S	116.46			33.29	83.62			−2.72	60.70			−3.51
C, C	−117.45			−32.30	−128.86			−39.85	−146.32			−49.58
C, C (2)	−21.87			−2.15	−17.08			0.07	−21.24			−0.80
C, N	−76.92	11.91		−4.27	−74.76	8.11		0.61	−108.24	26.37		−15.58
N, C	−1.13	−8.03	3.63	−9.18	−1.77	−9.32	2.59	−10.06	−1.59	−8.53	2.96	−10.07
N, C (2)				−25.86				−38.62				−35.84
N, C (3)	11.70			−13.63	66.30			32.25	5.30			−14.08
O, C	−67.09		10.89	12.19	5.51		26.90	74.61	−48.71		13.97	31.13
O, N	−96.91	−2.23	49.69	24.76	41.97	4.03	94.26	141.57	−71.01	−23.14	60.91	44.08
O, O	15.73	42.95	6.90	48.71	7.99	58.10	−4.37	39.39	19.46	48.95	8.03	51.79
O, P	132.11			127.66	59.17			77.23	−5.80			11.69
S, P	378.86			242.80	155.83			93.88	228.35			160.07
S, S	−0.03			−1.25	−7.51			−13.18	−6.54			−3.07
	$\hat{\sigma}_{\text{MR}}^{(\gamma)}$	$\hat{\sigma}_{\text{MR}}^{(\beta^2)}$	$\hat{\sigma}_{\text{MR}}^{(\phi^2)}$	$\hat{\sigma}_{\text{MR}}^{(\psi^2)}$	$\hat{\sigma}_{\text{MR}}^{(\gamma)}$	$\hat{\sigma}_{\text{MR}}^{(\beta^2)}$	$\hat{\sigma}_{\text{MR}}^{(\phi^2)}$	$\hat{\sigma}_{\text{MR}}^{(\psi^2)}$	$\hat{\sigma}_{\text{MR}}^{(\gamma)}$	$\hat{\sigma}_{\text{MR}}^{(\beta^2)}$	$\hat{\sigma}_{\text{MR}}^{(\phi^2)}$	$\hat{\sigma}_{\text{MR}}^{(\psi^2)}$
	0.1634	0.01	−1.13	−1.35	0.1503	0.12	−1.02	−1.44	0.1456	0.42	−0.98	−1.62

TABLE IV.

Performance of SM5CR Models by Solute Functional Class.

Solute Class	Number of Solute[a]	Number of Solvent Classes[b]	Number of Data Points[c]	SM5CR/AM1 MSE[d]	SM5CR/AM1 MUE[e]	SM5CR/PM3 MSE[d]	SM5CR/PM3 MUE[e]	SM5CR/MNDO/d MSE[d]	SM5CR/MNDO/d MUE[e]
Unbranched alkanes	9	19	84	0.28	0.60	0.29	0.62	0.23	0.61
Branched alkanes	5	3	12	0.04	0.58	−0.08	0.63	−0.23	0.70
Cycloalkanes	5	6	18	0.31	0.53	0.22	0.52	0.07	0.51
Alkenes	9	4	27	0.02	0.38	−0.05	0.37	−0.05	0.36
Alkynes	5	3	14	0.01	0.18	0.03	0.17	0.04	0.18
Arenes	9	19	134	−0.39	0.58	−0.45	0.60	−0.32	0.57
Alcohols	17	19	385	−0.02	0.61	−0.02	0.56	0.00	0.58
Ethers	12	19	93	−0.18	0.77	−0.02	0.60	−0.14	0.69
Aldehydes	7	8	38	0.08	0.81	−0.01	0.80	−0.05	0.81
Ketones	12	18	203	−0.24	0.54	−0.24	0.53	−0.20	0.49
Carboxylic acids	5	14	124	0.11	0.70	0.09	0.69	0.10	0.69
Esters	14	8	249	−0.07	0.63	−0.06	0.58	−0.07	0.61
Non-halo bifunctional compounds	5	8	28	0.61	1.24	0.67	1.24	0.43	1.13
Inorganic compounds	2	9	22	−0.03	0.66	−0.03	0.67	−0.03	0.64
Aliphatic amines	15	10	168	0.04	0.34	0.03	0.32	0.00	0.31
Aromatic amines	27	17	106	−0.03	0.46	0.02	0.47	0.04	0.50
Nitriles	4	6	22	0.00	0.34	0.00	0.37	0.00	0.42
Nitrohydrocarbons	6	8	38	0.02	0.59	−0.01	0.39	0.00	0.34
Amides & ureas, HCNO	9	13	25	0.53	0.95	0.59	0.90	0.67	1.01
Bifunctional HCN and HCNO	6	3	11	−0.27	0.45	−0.24	0.58	−0.38	0.60
Inorganic HCN	2	8	15	−0.09	0.56	−0.26	0.60	−0.10	0.53
Thiols	4	5	14	0.26	0.33	0.21	0.30	0.25	0.39
Sulfides	6	6	23	−0.07	0.62	−0.02	0.67	−0.03	0.50
Disulfides	2	3	5	0.00	0.22	−0.01	0.24	0.01	0.25
Fluorinated hydrocarbons	9	5	19	−0.25	0.70	−0.32	0.75	−0.60	1.16
Chloroalkanes	13	5	35	−0.56	0.69	−0.29	0.44	−0.52	0.70
Chloroalkenes	5	4	16	0.21	0.36	0.40	0.40	0.24	0.42
Chloroarenes	8	6	37	−0.25	0.47	−0.21	0.47	−0.49	0.62
Brominated hydrocarbons	14	6	50	−0.38	0.49	−0.47	0.56	−0.54	0.63
Iodinated hydrocarbons	9	6	28	−0.02	0.24	−0.09	0.46	−0.07	0.32

TABLE IV.
(Continued)

Solute Class	Number of			SM5CR/AM1		SM5CR/PM3		SM5CR/MNDO/d	
	Solute[a]	Solvent Classes[b]	Data Points[c]	MSE[d]	MUE[e]	MSE[d]	MUE[e]	MSE[d]	MUE[e]
Amides & ureas, halogen containing	27	1	16	−0.77	1.00	−0.63	0.96	−0.73	0.99
Aromatic sulfur compounds	17	1	17	−0.08	0.48	−0.05	0.40	−0.09	0.51
Other halo compounds	27	10	81	0.33	0.94	0.32	0.76	0.36	0.96
P, H, C, and O compounds	6	7	29	−0.40	1.76	−0.27	1.26	−0.39	1.44
Other P compounds	11	12	31	−0.10	1.90	−0.13	1.58	0.04	1.34
Total	332	19	2217	−0.05	0.63	−0.05	0.59	−0.06	0.61

New free energy of transfer data are distributed among the solute classes.
[a] Number of solutes in the given class.
[b] Number of solvent classes that contain data in the given solute class.
[c] Total number of solute–solvent data points in the solute class.
[d] Mean signed error (kcal/mol).
[e] Mean unsigned error (kcal/mol).

FIGURE 1. The mean absolute error versus the dielectric constant for each solvent in the training set for SM5CR/AM1.

screening energy calculated by COSMO should be less than $0.5\varepsilon^{-1}$ for solvents of dielectric constant ε. Hence, one might worry whether the relative error in the solvation free energies calculated by SM5CR may become large for solvents with small dielectric constants. This possibility is explored in Figure 1 in which the mean unsigned error for each solvent is presented graphically versus the dielectric constant. This plot indicates that, after parameterization of the atomic surface tensions, errors are not systematically larger for solvents with small dielectric constants than for those with large dielectric constants. In addition, Figure 2 divides the 91 solvents in the

FIGURE 2. The mean absolute error divided into groups of solvents with a similar dielectric constant ε for SM5CR/AM1.

TABLE V.
Performance of SM5CR Models by Solvent Functional Class.

	Number of			SM5CR/AM1		SM5CR/PM3		SM5CR/MNDO/d	
Solvent Class	Solvents[a]	Solute Classes[b]	Data Points[c]	MSE[d]	MUE[e]	MSE[d]	MUE[e]	MSE[d]	MUE[e]
Aqueous	1	31	251	−0.02	0.48	−0.01	0.47	−0.06	0.59
Alkanes	11	30	478	0.05	0.48	0.02	0.45	0.02	0.47
Cycloalkanes	2	24	106	0.24	0.50	0.17	0.47	0.15	0.48
Arenes	12	16	256	0.34	0.52	0.30	0.50	0.28	0.50
Aliphatic alcohols	12	31	299	−0.16	0.70	−0.14	0.66	−0.19	0.68
Aromatic alcohols	2	7	12	0.27	0.89	0.24	0.87	0.25	0.90
Ketones	4	10	35	−0.43	0.75	−0.34	0.70	−0.32	0.72
Esters	2	8	36	0.33	0.62	0.38	0.64	0.41	0.67
Aliphatic ethers	4	19	99	−0.25	0.64	−0.17	0.62	−0.19	0.61
Aromatic ethers	3	5	15	−0.67	0.86	−0.66	0.85	−0.66	0.82
Amines	2	6	12	0.45	0.56	0.32	0.55	0.27	0.65
Pyridines	4	5	15	−0.46	0.66	−0.40	0.62	−0.41	0.62
Nitriles	2	5	10	−1.10	1.22	−0.87	1.03	−0.81	1.07
Nitrohydrocarbons	4	8	27	−0.40	0.69	−0.26	0.79	−0.20	0.83
Tertiary amides	2	5	10	−0.20	0.60	−0.10	0.47	−0.12	0.54
Haloaliphatics	12	27	269	−0.23	0.68	−0.20	0.63	−0.20	0.64
Haloaromatics	6	11	106	−0.59	0.78	−0.50	0.68	−0.45	0.65
Miscellaneous acidic	3	5	15	0.21	0.62	0.21	0.58	0.20	0.59
Miscellaneous basic	4	12	39	0.31	0.71	0.22	0.62	0.23	0.66
Transfer free energies	7	5	67	−0.18	0.63	−0.14	0.61	−0.15	0.66
Total	91	33	2157	−0.05	0.59	−0.04	0.56	−0.05	0.59

Note that the new transfer free energy data are listed separately. Solutes containing phosphorus are not included in this table.
[a] Number of solvents in the given class.
[b] Number of solute classes that contain data in the given solvent class.
[c] Total number of solute–solvent data points in the solvent class.
[d] Mean signed error (kcal/mol).
[e] Mean unsigned error (kcal/mol).

TABLE VI.
Solvation Free Energies (kcal/mol) for Selected Neutral Solutes in Various Solvents.

Solute	Model	Solvent						
		Water	Hexadecane	Cyclohexane	Benzene	1-Octanol	Diethylether	Chloroform
n-Octane								
	SM5CR/AM1	2.08	−3.97	−4.28	−5.26	−4.04	−4.91	−4.88
	SM5CR/PM3	2.05	−4.02	−4.30	−5.23	−3.95	−4.76	−4.82
	SM5CR/MNDO/d	1.99	−4.14	−4.41	−5.35	−3.98	−4.74	−4.93
	Experimental	2.89	−5.02	−5.63	−5.35	−4.18	−5.62	−5.25
Benzene								
	SM5CR/AM1	−1.42	−4.62	−4.82	−5.60	−4.59	−5.49	−5.67
	SM5CR/PM3	−1.57	−4.79	−4.97	−5.84	−4.77	−5.60	−5.60
	SM5CR/MNDO/d	−1.54	−4.84	−5.01	−5.96	−4.81	−5.52	−5.44
	Experimental	−0.87	−3.80	−4.19	−4.55	−3.72	−4.21	−4.64
Toluene								
	SM5CR/AM1	−1.18	−4.81	−5.03	−5.88	−4.91	−5.80	−5.95
	SM5CR/PM3	−1.23	−4.93	−5.12	−6.04	−4.98	−5.81	−5.82
	SM5CR/MNDO/d	−1.05	−4.90	−5.10	−6.07	−4.87	−5.61	−5.56
	Experimental	−0.89	−4.54	−4.90	−5.32	−4.55	−5.23	−5.48
1-Butanol								
	SM5CR/AM1	−4.86	−3.15	−3.36	−4.25	−6.13	−5.86	−5.58
	SM5CR/PM3	−4.99	−3.33	−3.51	−4.41	−6.24	−5.80	−5.62
	SM5CR/MNDO/d	−5.03	−3.34	−3.52	−4.42	−6.29	−5.85	−5.68
	Experimental	−4.72	−3.55	−3.52	−4.45	−5.71	−5.69	−5.28
Phenol								
	SM5CR/AM1	−6.63	−5.12	−5.30	−6.27	−7.69	−7.69	−7.71
	SM5CR/PM3	−6.40	−5.13	−5.29	−6.32	−7.65	−7.51	−7.41
	SM5CR/MNDO/d	−6.50	−5.21	−5.37	−6.50	−7.79	−7.57	−7.32
	Experimental	−6.62	−5.14	−5.57	−7.12	−8.69	−8.75	−7.14
1,4-Dioxane								
	SM5CR/AM1	−5.06	−3.03	−3.22	−4.01	−5.52	−6.01	−5.53
	SM5CR/PM3	−5.12	−3.39	−3.56	−4.35	−5.26	−5.69	−5.38
	SM5CR/MNDO/d	−5.17	−3.22	−3.38	−4.14	−5.42	−5.86	−5.55
	Experimental	−5.05	−3.82	−4.17	−5.21	−4.89	−4.67	−6.21
Butanone								
	SM5CR/AM1	−3.69	−3.42	−3.61	−4.40	−5.40	−5.74	−5.61
	SM5CR/PM3	−3.65	−3.49	−3.67	−4.38	−5.47	−5.58	−5.71
	SM5CR/MNDO/d	−3.48	−3.44	−3.61	−4.37	−5.26	−5.51	−5.50
	Experimental	−3.64	−3.12	−3.48	−4.46	−3.78	−4.09	−5.43

TABLE VI.
(Continued)

Solute	Model	Solvent						
		Water	Hexadecane	Cyclohexane	Benzene	1-Octanol	Diethylether	Chloroform
Propanoic acid								
	SM5CR/AM1	−6.06	−3.17	−3.33	−4.13	−6.45	−6.74	−6.24
	SM5CR/PM3	−6.04	−3.17	−3.32	−4.10	−6.52	−6.70	−6.23
	SM5CR/MNDO/d	−6.06	−3.20	−3.34	−4.14	−6.55	−6.72	−6.28
	Experimental	−6.47	−3.12	−3.78	−4.75	−6.86	−6.75	−5.37
Butylamine								
	SM5CR/AM1	−3.55	−3.08	−3.30	−4.05	−5.02	−4.64	−4.93
	SM5CR/PM3	−3.56	−3.21	−3.42	−4.19	−5.01	−4.46	−4.78
	SM5CR/MNDO/d	−3.66	−3.20	−3.40	−4.14	−5.06	−4.49	−4.92
	Experimental	−4.29	−3.57	−3.72	−4.33	−5.35	−4.44	−5.35
Pyridine								
	SM5CR/AM1	−4.78	−4.65	−4.83	−5.47	−5.87	−5.65	−6.60
	SM5CR/PM3	−4.97	−4.84	−5.00	−5.68	−6.03	−5.57	−6.58
	SM5CR/MNDO/d	−5.01	−4.78	−4.94	−5.68	−6.10	−5.63	−6.47
	Experimental	−4.70	−4.10	−4.30	−5.28	−5.34	−4.81	−6.45
Aniline								
	SM5CR/AM1	−5.58	−5.30	−5.48	−6.43	−7.28	−7.34	−7.53
	SM5CR/PM3	−5.05	−5.23	−5.41	−6.44	−6.85	−6.88	−6.86
	SM5CR/MNDO/d	−5.18	−5.28	−5.45	−6.53	−7.04	−6.94	−6.88
	Experimental	−5.49	−5.44	−5.52	−6.88	−6.71	−6.51	−7.34
Nitrobenzene								
	SM5CR/AM1	−3.26	−5.48	−5.66	−6.17	−5.24	−7.13	−7.88
	SM5CR/PM3	−4.24	−5.84	−5.99	−6.31	−6.09	−7.76	−9.05
	SM5CR/MNDO/d	−4.35	−5.91	−6.05	−6.62	−6.27	−7.46	−8.48
	Experimental	−4.12	−6.22	−6.62	−7.60	−6.63	−6.85	−7.78
Thiophene								
	SM5CR/AM1	−0.48	−3.72	−3.92	−4.33	−2.88	−3.88	−4.75
	SM5CR/PM3	−0.05	−3.56	−3.74	−4.19	−2.81	−3.84	−4.48
	SM5CR/MNDO/d	−1.40	−4.00	−4.16	−4.78	−3.81	−4.13	−4.85
	Experimental	−1.42	−4.01	n/a	n/a	−3.89	n/a	−5.83
Chlorobenzene								
	SM5CR/AM1	−1.67	−5.09	−5.30	−6.18	−5.34	−6.11	−6.30
	SM5CR/PM3	−1.70	−5.23	−5.42	−6.36	−5.35	−6.08	−6.15
	SM5CR/MNDO/d	−1.97	−5.37	−5.55	−6.57	−5.69	−6.25	−6.27
	Experimental	−1.12	−4.99	−5.10	n/a	−5.00	−5.42	−5.45

TABLE VII.

Electrostatic and First-Solvation-Shell Contributions to Solvation Free Energies (kcal/mol) for Selected Neutral Solutes in Water and Chloroform.

Solute	Model	Water			Chloroform		
		ΔG_{EP}	G_{CDS}	ΔG_S^0	ΔG_{EP}	G_{CDS}	ΔG_S^0
n-Octane							
	SM5CR/AM1	−0.88	2.96	2.08	−0.69	−4.19	−4.88
	SM5CR/PM3	−0.38	2.43	2.05	−0.30	−4.52	−4.82
	SM5CR/MNDO/d	−0.01	2.00	1.99	0.00	−4.92	−4.93
	Experimental			2.89			−5.25
Benzene							
	SM5CR/AM1	−2.84	1.42	−1.42	−2.19	−3.48	−5.67
	SM5CR/PM3	−1.84	0.27	−1.57	−1.41	−4.19	−5.60
	SM5CR/MNDO/d	−0.62	−0.92	−1.54	−0.48	−4.96	−5.44
	Experimental			−0.87			−4.64
Toluene							
	SM5CR/AM1	−2.93	1.75	−1.18	−2.26	−3.70	−5.95
	SM5CR/PM3	−1.88	0.65	−1.23	−1.45	−4.37	−5.82
	SM5CR/MNDO/d	−0.59	−0.46	−1.05	−0.46	−5.10	−5.56
	Experimental			−0.89			−5.48
1-Butanol							
	SM5CR/AM1	−6.26	1.40	−4.86	−4.82	−0.77	−5.58
	SM5CR/PM3	−5.48	0.49	−4.99	−4.20	−1.42	−5.62
	SM5CR/MNDO/d	−5.52	0.49	−5.03	−4.26	−1.43	−5.68
	Experimental			−4.72			−5.28
Phenol							
	SM5CR/AM1	−7.62	0.99	−6.63	−5.84	−1.87	−7.71
	SM5CR/PM3	−6.21	−0.18	−6.40	−4.75	−2.67	−7.41
	SM5CR/MNDO/d	−5.20	−1.30	−6.50	−4.00	−3.32	−7.32
	Experimental			−6.62			−7.14
1,4-Dioxane							
	SM5CR/AM1	−7.98	2.92	−5.06	−6.16	0.62	−5.53
	SM5CR/PM3	−5.88	0.76	−5.12	−4.54	−0.84	−5.38
	SM5CR/MNDO/d	−7.10	1.93	−5.17	−5.51	−0.04	−5.55
	Experimental			−5.05			−6.21
Butanone							
	SM5CR/AM1	−6.38	2.69	−3.69	−4.75	−0.86	−5.61
	SM5CR/PM3	−6.17	2.51	−3.65	−4.59	−1.13	−5.71
	SM5CR/MNDO/d	−5.56	2.08	−3.48	−4.16	−1.34	−5.50
	Experimental			−3.64			−5.43

TABLE VII.
(Continued)

Solute	Model	Water			Chloroform		
		ΔG_{EP}	G_{CDS}	ΔG_S^0	ΔG_{EP}	G_{CDS}	ΔG_S^0
Propanoic acid							
	SM5CR/AM1	−10.14	4.08	−6.06	−7.76	1.52	−6.24
	SM5CR/PM3	−9.91	3.87	−6.04	−7.57	1.34	−6.23
	SM5CR/MNDO/d	−9.81	3.75	−6.06	−7.53	1.26	−6.28
	Experimental			−6.47			−5.37
Butylamine							
	SM5CR/AM1	−3.64	0.09	−3.55	−2.86	−2.07	−4.93
	SM5CR/PM3	−2.26	−1.30	−3.56	−1.77	−3.01	−4.78
	SM5CR/MNDO/d	−2.68	−0.99	−3.66	−2.11	−2.82	−4.93
	Experimental			−4.29			−5.35
Pyridine							
	SM5CR/AM1	−4.76	−0.02	−4.78	−3.66	−2.94	−6.60
	SM5CR/PM3	−3.58	−1.39	−4.97	−2.74	−3.84	−6.58
	SM5CR/MNDO/d	−3.18	−1.83	−5.01	−2.45	−4.02	−6.47
	Experimental			−4.70			−6.45
Aniline							
	SM5CR/AM1	−6.23	0.65	−5.58	−4.84	−2.69	−7.53
	SM5CR/PM3	−3.76	−1.29	−5.05	−2.91	−3.95	−6.86
	SM5CR/MNDO/d	−3.31	−1.88	−5.18	−2.58	−4.30	−6.88
	Experimental			−5.49			−7.34
Nitrobenzene							
	SM5CR/AM1	−9.71	6.44	−3.26	−7.43	−0.45	−7.88
	SM5CR/PM3	−12.73	8.49	−4.24	−9.88	0.84	−9.05
	SM5CR/MNDO/d	−9.19	4.84	−4.35	−7.04	−1.44	−8.48
	Experimental			−4.12			−7.78
Thiophene							
	SM5CR/AM1	−2.91	2.43	−0.48	−2.25	−2.50	−4.75
	SM5CR/PM3	−2.54	2.49	−0.05	−1.95	−2.53	−4.48
	SM5CR/MNDO/d	−0.99	−0.40	−1.40	−0.77	−4.08	−4.85
	Experimental			−1.42			−5.83
Chlorobenzene							
	SM5CR/AM1	−3.12	1.45	−1.67	−2.39	−3.91	−6.30
	SM5CR/PM3	−1.95	0.25	−1.70	−1.49	−4.66	−6.15
	SM5CR/MNDO/d	−1.38	−0.59	−1.97	−1.05	−5.22	−6.27
	Experimental			−1.12			−5.45

TABLE VIII.

Solvation Free Energies of Ions, Presented with Electrostatic and Nonelectrostatic Contributions (kcal/mol), Calculated by SM5CR Models.

	SM5CR/AM1			SM5CR/PM3			SM5CR/MNDO/d			Experiment
Solute	ΔG_{EP}	G_{CDS}	ΔG_S^0	ΔG_{EP}	G_{CDS}	ΔG_S^0	ΔG_{EP}	G_{CDS}	ΔG_S^0	ΔG_S^0
HC_2^-	−83.4	3.3	−80.1	−85.7	2.6	−83.1	−82.6	1.7	−80.9	−73
$CH_3OH_2^+$	−90.4	3.8	−86.7	−90.1	4.3	−85.8	−89.5	2.9	−86.6	−87
$(CH_3)_2OH^+$	−74.3	1.7	−72.5	−74.4	1.9	−72.5	−73.7	1.3	−72.4	−70
$CH_3CH_2OH_2^+$	−84.4	3.4	−81.0	−84.0	3.9	−80.0	−84.3	2.4	−81.9	−81
$CH_3C(OH)CH_3^+$	−71.7	1.5	−70.2	−71.2	1.8	−69.3	−72.2	0.9	−71.4	−64
H_3O^+	−112.6	11.3	−101.3	−110.5	11.7	−98.8	−110.9	9.3	−101.6	−105
CH_3O^-	−90.3	2.4	−88.0	−92.2	1.7	−90.5	−89.0	2.1	−86.9	−98
$CH_3CO_2^-$	−82.2	6.5	−75.7	−83.5	5.9	−77.7	−82.7	6.5	−76.2	−77
$CH_3COCH_2^-$	−77.4	2.9	−74.5	−78.2	2.5	−75.7	−76.3	2.2	−74.1	−81
$C_6H_5O^-$	−71.3	2.2	−69.1	−71.7	1.1	−70.6	−68.9	0.2	−68.7	−75
$C_6H_5CH_2^-$	−61.5	2.1	−59.4	−60.7	0.8	−59.9	−57.6	−0.4	−58.0	−59
OH^-	−114.0	0.0	−114.0	−115.4	−2.2	−117.6	−113.2	−0.9	−114.1	−110
HO_2^-	−101.1	4.5	−96.6	−102.5	0.1	−102.4	−99.7	3.5	−96.2	−101
O_2^-	−93.1	6.5	−86.6	−93.4	0.1	−93.3	−92.9	5.3	−87.5	−87
$CH_3NH_3^+$	−76.7	1.1	−75.7	−76.1	−0.1	−76.2	−76.3	−0.2	−76.5	−73
$HC(OH)NH_2^+$	−79.2	0.3	−78.9	−76.0	−0.1	−76.0	−78.4	−0.7	−79.0	−78
CH_3CNH^+	−69.7	1.0	−68.7	−69.8	0.8	−69.0	−69.9	−0.1	−70.0	−69
$CH_3C(OH)NH_2^+$	−71.0	0.4	−70.6	−69.0	−0.4	−69.4	−70.4	−0.7	−71.2	−70
$(CH_3)_2NH_2^+$	−70.0	2.0	−68.0	−69.8	1.3	−68.6	−69.5	1.0	−68.5	−66
$(CH_3)_3NH^+$	−64.2	2.6	−61.7	−64.3	2.2	−62.2	−63.5	1.8	−61.7	−59
$ImidazoleH^+$	−66.6	1.4	−65.2	−63.8	0.3	−63.5	−64.4	−0.6	−65.0	−64
$C_5H_5NH^+$	−62.6	1.5	−61.1	−61.6	0.3	−61.3	−60.7	−0.7	−61.4	−58
$C_6H_5NH_3^+$	−69.9	1.1	−68.7	−66.8	−0.8	−67.6	−66.8	−1.8	−68.6	−68
NH_4^+	−85.1	0.0	−85.1	−83.4	−1.8	−85.2	−84.6	−1.7	−86.3	−81

TABLE VIII.
(Continued)

Solute	SM5CR/AM1			SM5CR/PM3			SM5CR/MNDO/d			Experiment
	ΔG_{EP}	G_{CDS}	ΔG_S^0	ΔG_{EP}	G_{CDS}	ΔG_S^0	ΔG_{EP}	G_{CDS}	ΔG_S^0	ΔG_S^0
CN^-	−78.5	2.9	−75.6	−77.9	3.9	−74.0	−77.9	2.9	−75.0	−75
CH_2CN^-	−69.7	0.6	−69.0	−69.2	1.1	−68.1	−68.8	0.1	−68.8	−75
NH_2^-	−84.1	−0.6	−84.7	−84.7	−2.6	−87.2	−83.8	−1.0	−84.8	−95
NO_2^-	−80.7	7.2	−73.5	−81.3	10.5	−70.8	−80.8	8.0	−72.8	−73
NO_3^-	−76.6	12.2	−64.5	−82.1	17.3	−64.8	−77.1	13.0	−64.1	−66
N_3^-	−66.5	−1.8	−68.3	−69.7	−4.7	−74.3	−66.6	−0.5	−67.0	−74
$CH_3SH_2^+$	−72.9	4.7	−68.2	−73.9	3.3	−70.6	−70.5	1.5	−69.0	−74
F^-	−109.7	2.7	−107.0	−109.7	2.7	−107.0	−109.7	3.8	−105.9	−107
HS^-	−81.1	3.2	−77.9	−79.3	3.6	−75.7	−78.3	−0.9	−79.2	−76
CH_3S^-	−78.9	2.2	−76.7	−77.9	3.2	−74.7	−76.0	−0.3	−76.3	−76
$CH_3CH_2S^-$	−77.3	2.2	−75.1	−76.3	3.0	−73.3	−74.5	−0.3	−74.7	−74
n-$C_3H_7S^-$	−77.2	2.4	−74.8	−76.0	3.2	−72.8	−74.5	−0.1	−74.6	−76
$C_6H_5S^-$	−67.8	2.1	−65.7	−67.7	2.1	−65.6	−64.3	−1.9	−66.2	−65
$(CH_3)_2SH^+$	−65.7	3.8	−61.9	−66.2	3.2	−63.0	−64.4	1.8	−62.6	−61
$CHF_2CO_2^-$	−75.0	7.5	−67.4	−76.6	7.1	−69.5	−76.5	8.6	−67.9	−70
Cl^-	−77.6	0.6	−77.0	−77.6	0.2	−77.4	−77.6	0.5	−77.1	−78
$CHCl_2CO_2^-$	−67.2	5.7	−61.5	−69.5	5.1	−64.5	−67.4	5.9	−61.4	−66
Br^-	−71.6	0.4	−71.3	−71.6	0.2	−71.4	−71.6	0.4	−71.2	−72
I^-	−62.4	−0.3	−62.7	−62.4	0.8	−61.7	−62.4	0.2	−62.2	−63
PH_2^-	−68.4	4.1	−64.3	−67.1	0.4	−66.7	−65.0	0.0	−65.0	−67
PH_4^+	−64.9	4.0	−60.9	−65.9	2.9	−63.1	−65.0	2.8	−62.2	−73
$CH_3PH_3^+$	−61.8	3.7	−58.1	−63.2	3.1	−60.1	−61.7	3.0	−58.7	−63
$(CH_3)_2PH_2^+$	−60.0	3.4	−56.5	−60.8	3.2	−57.7	−59.1	3.0	−56.1	−57
$(CH_3)_3PH^+$	−58.8	3.3	−55.5	−58.6	3.0	−55.6	−56.8	2.9	−54.0	−53
$H_2PO_4^-$	−85.9	24.6	−61.3	−81.1	11.4	−69.8	−77.4	8.0	−69.4	−68
Mean unsigned error			3.0			2.7			3.1	
Mean signed error			1.1			0.3			0.8	

training set into groups of similar dielectric constants. This figure further indicates that the SM5CR models are equally accurate for solvents with a wide range of dielectric constants. Figures 1 and 2 were generated using SM5CR/AM1, although the other two parameterizations would demonstrate similar behavior. The slightly better performance realized by using $(\varepsilon - 1)/\varepsilon$ instead of $(\varepsilon - 1)/(\varepsilon + 0.5)$ (*vide supra*) comes from all ranges of ε, not particularly from the lowest ε range.

The solvation free energies of ions are generally dominated by electrostatic effects, and thus the performance of the SM5CR models for ionic compounds was studied to gauge the accuracy of our COSMO electrostatic calculations. Note that for ionic compounds the typical uncertainty in the experimental solvation free energy is estimated to be 5 kcal/mol. Table VIII confirms that the overall solvation free energy of ions is dominated by the ΔG_{EP} contribution and that the G_{CDS} term provides only a relatively small correction. Most importantly, trends in ΔG_{EP} match those in the experimental values of ΔG^0_S. The overall mean signed and unsigned errors for the 49 ions in Table VIII are comparable to other SM*x* models.[4, 6, 8]

It might be useful to add a few qualitative remarks on the comparison between the COSMO algorithm used here and the GB model used in the previous Minnesota solvation models. In principle, the approximation of dielectric media as conductor-like appears accurate enough that COSMO should often provide a more faithful solution to the electrostatic problem than does the GB method. Furthermore, it includes the atomic dipoles (due to the one-center off-diagonal elements of the solute density matrix) in the charge density, whereas the GB model is restricted to distributed atomic monopoles. There are, however, some competing advantages to the GB model. GB probably agrees with a numerical solution of the full electrostatic problem (the Poisson equation) about as well as the Poisson model mimics the true physical situation, but it yields a very "smooth" result as a function of geometry (less numerical noise than COSMO or Poisson solvers like PCM) and our current gradient algorithms for GB are efficient and essentially devoid of methodological noise.[9] (In particular they avoid the numerical noise from the tesselation and the approximated analytical gradient.) The GB methods are also appealing in that they do not require truncating the solute charge distribution at the cavity–solvent boundary, which leads to a disconcerting outlying charge error in all methods (including COSMO) that employ apparent surface charges on the solute cavity surface.[10] Furthermore, the GB method allows us to use[1, 2, 5–9] class IV charges,[42] which are much more accurate than the wave functions or density matrix from which they derive. There is no evidence that one approach to electrostatics yields consistently more accurate values than the other, especially when the electrostatics are augmented by CDS terms. All things considered, we view the COSMO and GB methods as complementary with compensating advantages and disadvantages.[10] We do note, however, the historically important fact that the current COSMO effort is also our most recent effort and because we improved the training set in this study the current SM5CR model may be more robust than our current GB SM*x* parameterizations for the kinds of molecules that dominate the additions to the training set (i.e., mononitrogen heterocycles, amides, ureas, sulfur compounds, and phosphonofluoridates). Furthermore, because we returned to using a nonzero solvent radius, potentials of mean force for interacting solute pairs may be more physical than in models with zero solute radii.

Although analytic gradients have not been coded for the present implementation of the SM5CR model in AMPAC, the formalisms for analytic gradients of COSMO electrostatics[17] and SM5.42 surface tensions[9] are both available. If, in later work, geometries are optimized in liquid-phase solution (using either analytic or numerical gradients), the model should be denoted SM5C. In SM5C calculations the ΔG_{EP} would be replaced by ΔG_{ENP}, which also includes the relaxation of the nuclear (N) framework in the liquid environment. In rigid models the ΔG_{ENP} becomes ΔG_{EP} because the solute nuclear repulsion does not change. The label SM5CR refers to calculations employing gas-phase geometries. We do not consider it essential that the gas-phase geometries be calculated by a particular method; any reasonable gas-phase geometries may be used.

Concluding Remarks

The COSMO method for the electrostatic part of free energies of solvation is combined with the SM5 functional forms for nonelectrostatic effects.[43] It is established that the accuracy and precision of COSMO calculations can be improved by adjusting the default values of two parameters that affect the numerical methods of the model. Additions to the standard SM*x* training set increase the total number

TABLE S-I.
Solvent Properties Used to Parameterize the SM5CR Model.

Compound Name	n	α	β	γ[a]	ε	ϕ	ψ
Acetic acid	1.372	0.61	0.44	39.01	6.25	0.00	0.00
Acetonitrile	1.3442	0.07	0.32	41.25	37.5	0.00	0.00
Acetophenone	1.5372	0	0.48	56.19	17.44	0.67	0.00
Aniline	1.5863	0.26	0.41	60.62	6.89	0.86	0.00
Anisole	1.5174	0	0.29	50.52	4.22	0.75	0.00
Benzene	1.5011	0	0.14	40.62	2.27	1.00	0.00
Benzonitrile	1.5289	0	0.33	44.32	25.59	0.75	0.00
Benzyl alcohol	1.5396	0.33	0.56	50.09	11.92	0.75	0.00
Bromobenzene	1.5597	0	0.09	50.72	5.4	0.86	0.14
Bromoethane	1.4239	0	0.12	34	9.02	0.00	0.33
Bromoform	1.6005	0.15	0.06	64.58	4.25	0.00	0.75
1-Bromooctane	1.4524	0	0.12	41.28	5.02	0.00	0.11
1-Butanol	1.3993	0.37	0.48	35.88	17.33	0.00	0.00
2-Butanol	1.3978	0.33	0.56	32.44	15.94	0.00	0.00
Butanone	1.3788	0	0.51	34.5	18.25	0.00	0.00
Butyl acetate	1.3941	0	0.45	35.69	4.99	0.00	0.00
n-Butylbenzene	1.4898	0	0.15	41.34	2.36	0.60	0.00
sec-Butylbenzene	1.4895	0	0.16	40.34	2.34	0.60	0.00
tert-Butylbenzene	1.4927	0	0.16	39.78	2.34	0.60	0.00
Carbon disulfide	1.6319	0	0.07	45.45	2.61	0.00	0.00
Carbon tetrachloride	1.4601	0	0	38.04	2.23	0.00	0.80
Chlorobenzene	1.5241	0	0.07	47.48	5.7	0.86	0.14
Chloroform	1.4459	0.15	0.02	38.39	4.71	0.00	0.75
1-Chlorohexane	1.4199	0	0.1	37.03	5.95	0.00	0.14
m-Cresol	1.5438	0.57	0.34	51.37	12.7	0.75	0.00
Cyclohexane	1.4266	0	0	35.48	2.02	0.00	0.00
Cyclohexanone	1.4507	0	0.56	49.76	15.62	0.00	0.00
Decalin	1.4753	0	0	43.87	2.2	0.00	0.00
n-Decane	1.4102	0	0	33.64	1.99	0.00	0.00
1-Decanol	1.4372	0.37	0.48	41.03	7.53	0.00	0.00
1,2-Dibromoethane	1.542	0.1	0.1	56.92	4.93	0.00	0.50
Dibutyl ether	1.3992	0	0.45	32.3	3.05	0.00	0.00
o-Dichlorobenzene	1.5515	0	0.04	52.69	9.99	0.75	0.25
1,2-Dichloroethane	1.4448	0.1	0.11	45.86	10.19	0.00	0.50
Diethyl ether	1.3526	0	0.41	23.96	4.24	0.00	0.00
Diisopropyl ether	1.3679	0	0.41	24.86	3.81	0.00	0.00
N,*N*-Dimethylacetamide	1.438	0	0.78	47.71	37.78	0.00	0.00
N,*N*-Dimethylformamide	1.4305	0	0.74	50.66	37.22	0.00	0.00
2,6-Dimethylpyridine	1.4953	0	0.63	44.63	7.17	0.63	0.00
Dimethylsulfoxide	1.417	0	0.88	61.77	46.83	0.00	0.00
Diphenyl ether	1.5787	0	0.2	38.5	3.73	0.92	0.00
n-Dodecane	1.4216	0	0	35.85	2.01	0.00	0.00
Ethanol	1.3611	0.37	0.48	31.62	24.85	0.00	0.00
Ethoxybenzene	1.4959	0	0.32	46.26	4.18	0.67	0.00
Ethyl acetate	1.3723	0	0.45	33.67	5.99	0.00	0.00
Ethylbenzene	1.4959	0	0.15	41.38	2.43	0.75	0.00
Fluorobenzene	1.4684	0	0.1	38.37	5.47	0.86	0.14
1-Fluoroctane	1.3935	0	0.1	33.92	3.89	0.00	0.11
n-Heptane	1.3878	0	0	28.28	1.91	0.00	0.00
1-Heptanol	1.4249	0.37	0.48	38.17	11.32	0.00	0.00
n-Hexadecane	1.4345	0	0	38.93	2.06	0.00	0.00
n-Hexane	1.3749	0	0	25.75	1.88	0.00	0.00
1-Hexanol	1.4178	0.37	0.48	37.15	12.51	0.00	0.00

TABLE S-I.
(Continued)

Compound Name	n	α	β	γ[a]	ε	ϕ	ψ
Iodobenzene	1.62	0	0.12	55.72	4.55	0.86	0.00
1-Iodohexadecane	1.4806	0	0.15	46.47	3.53	0.00	0.00
Isobutanol	1.3955	0.37	0.48	32.44	16.78	0.00	0.00
Isopropanol	1.3776	0.33	0.56	30.12	19.26	0.00	0.00
Isopropylbenzene	1.4915	0	0.16	39.84	2.37	0.67	0.00
p-Isopropyltoluene	1.4909	0	0.19	38.34	2.23	0.60	0.00
2-Methoxyethanol	1.4024	0.3	0.84	44.39	17.11	0.00	0.00
Methylene chloride	1.4242	0.1	0.05	39.15	8.82	0.00	0.67
N-Methylformamide	1.4319	0.4	0.55	55.7	181.56	0.00	0.00
4-Methyl-2-pentanone	1.3962	0	0.51	33.35	12.89	0.00	0.00
2-Methylpyridine	1.4957	0	0.58	47.5	9.95	0.71	0.00
Nitrobenzene	1.5562	0	0.28	62.54	34.81	0.67	0.00
Nitroethane	1.3917	0.02	0.33	46.24	28.29	0.00	0.00
Nitromethane	1.3817	0.06	0.31	52.58	36.56	0.00	0.00
n-Nonane	1.4054	0	0	32.23	1.96	0.00	0.00
1-Nonanol	1.4333	0.37	0.48	40.14	8.6	0.00	0.00
n-Octane	1.3974	0	0	30.43	1.94	0.00	0.00
1-Octanol	1.4295	0.37	0.48	39.01	9.87	0.00	0.00
o-Nitrotoluene	1.545	0	0.27	59.13	25.67	0.60	0.00
n-Pentadecane	1.4315	0	0	38.34	2.03	0.00	0.00
n-Pentane	1.3575	0	0	22.29	1.84	0.00	0.00
1-Pentanol	1.4101	0.37	0.48	36.5	15.13	0.00	0.00
Perfluorobenzene	1.3777	0	0	31.66	2.03	0.50	0.50
1-Propanol	1.385	0.37	0.48	33.56	20.52	0.00	0.00
Pyridine	1.5095	0	0.52	52.62	12.98	0.83	0.00
Sulfolane[b]	1.4833	0	0.88	77.62	43.96	0.00	0.00
Tetrachloroethene	1.5053	0	0	46.63	2.27	0.00	0.67
Tetrahydrofuran	1.405	0	0.48	38	7.43	0.00	0.00
Tetralin	1.5413	0	0.19	47.74	2.77	0.60	0.00
Toluene	1.4961	0	0.14	40.2	2.38	0.86	0.00
Tributyl phosphate	1.4224	0	1.21	37.39	8.18	0.00	0.00
Triethylamine	1.401	0	0.79	29.1	2.38	0.00	0.00
1,2,4-Trimethylbenzene	1.5048	0	0.19	42.03	2.37	0.67	0.00
1,3,5-Trimethylbenzene[c]	1.4994	0	0.19	39.65	2.27	0.67	0.00
2,2,4-Trimethylpentane	1.4513	0	0	26.38	1.94	0.00	0.00
n-Undecane	1.4398	0	0	34.85	1.99	0.00	0.00
Xylenes (tech. mixture)	1.4995	0	0.16	41.38	2.39	0.75	0.00

[a] In cal mol^{-1} $Å^{-2}$.
[b] Tetrahydrothiophene-*S*,*S*-dioxide.
[c] Mesitylene.

TABLE S-II.

New Water–Solvent Partition Coefficient Data Added to Training Set, Including Experimental Value Source.

Source	Solute	Solvent	AM1			PM3	MNDO/d	Experimental
			$G^0_{S,water}$	$G^0_{S,solv}$	log *P*	log *P*	log *P*	log *P*
M	4-Aminopyridine	1-Octanol	−9.23	−8.77	−0.33	−0.31	−0.28	0.30
M	2-Cyanopyrrole	1-Octanol	−4.00	−6.84	2.08	2.02	2.04	1.13
M	3,5-Dimethylpyridine	Benzene	−4.06	−5.93	1.37	1.39	1.47	1.51
M	2,5-Dimethylpyrrole	1-Octanol	−4.39	−7.30	2.14	2.20	2.27	1.47
M	4-Ethylpyridine	1-Octanol	−4.32	−6.52	1.61	1.56	1.60	1.65
M	4-(2-Hydroxyethyl)pyridine	1-Octanol	−10.64	−10.43	−0.15	−0.17	−0.13	0.10
M	4-(Hydroxymethyl)pyridine	1-Octanol	−10.03	−9.63	−0.29	−0.28	−0.26	0.06
M	Isoquinoline	Cyclohexane	−6.05	−7.08	0.76	0.81	0.81	1.11
M	Isoquinoline	*n*-Heptane	−6.05	−7.19	0.84	0.86	0.85	1.01
M	Isoquinoline	1-Octanol	−6.05	−8.21	1.58	1.57	1.63	2.08
M	*N*-Methylpyrrole	1-Octanol	−3.53	−6.06	1.86	1.98	2.03	1.21
M	4-Nitropyridine	1-Octanol	−5.51	−5.58	0.05	−0.13	−0.06	0.33
M	*N*-Phenylpyrrole	1-Octanol	−3.91	−8.12	3.09	3.20	3.26	3.08
M	Pyrrole	Chloroform	−5.29	−6.52	0.90	0.78	0.85	0.91
M	Pyrrole	Cyclohexane	−5.29	−4.45	−0.61	−0.55	−0.20	−0.36
M	Pyrrole	1-Octanol	−5.29	−6.34	0.77	0.93	0.95	0.75
M	Quinoline	Chloroform	−5.40	−9.17	2.77	2.58	2.40	3.35
M	Quinoline	Cyclohexane	−5.40	−7.18	1.31	1.39	1.31	1.26
M	Quinoline	1-Octanol	−5.40	−8.12	1.99	2.00	2.04	2.04
M	Acetanilide	Benzene	−8.42	−8.69	0.20	0.27	0.48	0.22
M	Acetanilide	Chloroform	−8.42	−10.74	1.70	1.62	1.50	0.88
M	Acetanilide	Cyclohexane	−8.42	−7.49	−0.68	−0.64	−0.51	−1.41
M	Acetanilide	Ethyl ether	−8.42	−10.50	1.53	1.43	1.51	0.54
M	Acetanilide	*n*-Heptane	−8.42	−7.59	−0.61	−0.60	−0.48	−1.32
M	Acetanilide	*n*-Hexane	−8.42	−7.64	−0.57	−0.57	−0.46	−0.99
M	Acetanilide	1-Octanol	−8.42	−10.37	1.43	1.53	1.54	1.16
M	4-Cyanoacetanilide	1-Octanol	−11.30	−11.93	0.46	0.62	0.48	1.37
M	3-Methylacetanilide	1-Octanol	−8.19	−10.69	1.83	1.93	1.94	1.57
M	4-Methylacetanilide	1-Octanol	−8.17	−10.67	1.83	1.92	1.94	1.57
M	Phenylurea	Chloroform	−12.19	−11.67	−0.38	−0.55	−0.68	−0.63
M	Phenylurea	Ethyl ether	−12.19	−11.48	−0.52	−0.63	−0.60	−0.11
M	Phenylurea	*n*-Heptane	−12.19	−7.65	−3.33	−3.24	−3.27	−3.16
M	Phenylurea	1-Octanol	−12.19	−12.02	−0.13	−0.01	−0.06	0.83

TABLE S-II.
(Continued)

Source	Solute	Solvent	AM1			PM3	MNDO/d	Experimental
			$G^0_{S,water}$	$G^0_{S,solv}$	log *P*	log *P*	log *P*	log *P*
M	4-Bromoacetanilide	1-Octanol	−8.70	−11.55	2.09	2.19	2.23	2.11
M	3-Bromophenylurea	1-Octanol	−12.69	−13.35	0.49	0.61	0.59	2.08
M	4-Bromophenylurea	1-Octanol	−12.49	−13.21	0.53	0.65	0.63	1.98
M	3-Chlorophenylurea	1-Octanol	−12.44	−12.73	0.22	0.31	0.24	1.82
M	4-Chlorophenylurea	1-Octanol	−12.23	−12.59	0.26	0.33	0.28	1.80
M	2,6-Difluoroacetanilide	1-Octanol	−7.52	−9.61	1.53	1.58	1.49	0.69
M	1,1-Dimethyl-3-(4-bromo-phenyl)urea	1-Octanol	−9.69	−12.72	2.22	2.37	2.33	2.19
M	1,1-Dimethyl-3-(3-fluoro-phenyl)urea	1-Octanol	−9.11	−11.16	1.50	1.61	1.48	1.37
M	1,1-Dimethyl-3-(4-fluoro-phenyl)urea	1-Octanol	−9.13	−11.17	1.49	1.61	1.48	1.13
M	3-Fluoroacetanilide	1-Octanol	−8.08	−9.97	1.38	1.44	1.40	1.57
M	4-Fluoroacetanilide	1-Octanol	−8.02	−9.93	1.40	1.46	1.42	1.38
M	3-Fluorophenylurea	1-Octanol	−11.86	−11.62	−0.17	−0.09	−0.20	1.29
M	4-Fluorophenylurea	1-Octanol	−11.76	−11.55	−0.16	−0.08	−0.19	1.04
M	4-Iodoacetanilide	1-Octanol	−8.79	−11.76	2.18	2.48	2.32	2.58
M	3-Trifluoromethylacetanilide	1-Octanol	−7.43	−9.91	1.82	1.83	1.80	2.38
M	3-Trifluoromethylphenylurea	1-Octanol	−8.37	−11.03	1.95	2.02	1.90	2.31
M	4-Bromopyridine	1-Octanol	−4.85	−6.87	1.48	1.46	1.52	1.54
B	1,2-Benzodithiolan-3-one	1-Octanol	−3.18	−5.63	1.79	1.88	1.74	2.73
B	1,2-Benzodithiolane-3-thione	1-Octanol	−2.22	−6.37	3.04	2.95	3.00	3.57
B	4,5-Dimethyl-1,2-dithiolan-3-one	1-Octanol	−2.00	−4.44	1.79	1.87	1.73	1.73
B	4,5-Dimethyl-1,2-dithiolane-3-thione	1-Octanol	−1.65	−5.58	2.88	2.80	2.88	2.45
B	1,2-Dithiolan-3-one	1-Octanol	−3.49	−4.23	0.54	0.70	0.55	0.83
B	1,2-Dithiolane-3-thione	1-Octanol	−3.03	−5.64	1.91	1.90	1.97	1.65
B	4-Methyl-1,2-dithiolan-3-one	1-Octanol	−2.48	−4.16	1.24	1.34	1.18	1.33
B	5-Methyl-1,2-dithiolan-3-one	1-Octanol	−2.83	−4.39	1.14	1.28	1.14	1.26
B	4-Methyl-1,2-dithiolane-3-thione	1-Octanol	−2.00	−5.19	2.34	2.24	2.34	2.18
B	5-Methyl-1,2-dithiolane-3-thione	1-Octanol	−2.41	−5.85	2.53	2.54	2.56	1.87
B	5-Methyl-4-phenyl-1,2-dithiolan-3-one	1-Octanol	−3.26	−7.07	2.79	2.88	2.78	2.93
B	4-Methyl-5-phenyl-1,2-dithiolane-3-thione	1-Octanol	−2.33	−7.63	3.88	3.83	3.85	3.95
B	5-Methyl-4-phenyl-1,2-dithiolane-3-thione	1-Octanol	−2.87	−8.17	3.88	3.78	3.92	3.17
B	4-Phenyl-1,2-dithiolan-3-one	1-Octanol	−3.71	−6.94	2.37	2.47	2.35	2.64
B	5-Phenyl-1,2-dithiolan-3-one	1-Octanol	−3.74	−6.64	2.13	2.26	2.12	3.01
B	4-Phenyl-1,2-dithiolane-3-thione	1-Octanol	−3.34	−8.00	3.41	3.28	3.46	3.20

TABLE S-II.
(Continued)

Source	Solute	Solvent	AM1			PM3	MNDO/d	Experimental
			$G^0_{S,water}$	$G^0_{S,solv}$	log *P*	log *P*	log *P*	log *P*
B	5-Phenyl-1,2-dithiolane-3-thione	1-Octanol	−3.37	−8.16	3.51	3.57	3.52	3.67
M	Dimethylphenylphosphine	1-Octanol	−0.47	−3.88	2.50	2.48	2.52	2.57
M	Isopropyl methylphosphonofluoridate	Benzene	−8.77	−6.70	−1.52	−1.10	−0.32	0.32
M	Isopropyl methylphosphonofluoridate	Carbon tetrachloride	−8.77	−6.67	−1.54	−1.16	−0.28	−0.08
M	Isopropyl methylphosphonofluoridate	Chlorobenzene	−8.77	−10.80	1.49	1.22	1.25	0.29
M	Isopropyl methylphosphonofluoridate	Chloroform	−8.77	−10.58	1.32	1.03	1.21	1.49
M	Isopropyl methylphosphonofluoridate	1,2-Dibromoethane	−8.77	−8.77	0.00	−0.03	0.12	0.35
M	Isopropyl methylphosphonofluoridate	1,2-Dichloroethane	−8.77	−11.83	2.25	1.68	1.50	1.25
M	Isopropyl methylphosphonofluoridate	*n*-Heptane	−8.77	−5.72	−2.24	−1.88	−0.86	−0.70
M	Isopropyl methylphosphonofluoridate	Nitrobenzene	−8.77	−11.86	2.27	1.81	1.40	0.40

M, Medchem data base[36]; B, Bona et al.[37] Calculated values of log *P* are given, and SM5CR/AM1 values are partitioned into calculated $G^0_{S,water}$ and $G^0_{S,solv}$ values (kcal/mol).

of data points in the training set from 2135 solvation free energies data points to 2217 total data points for neutral solutes, which now include solvent–solvent partitioning data in addition to solvation free energy data. The end result is three new parameterizations, which yield an overall mean absolute error, averaged over the three parameterizations, of 0.61 kcal/mol. Although the overall mean absolute error is larger than that quoted for other SM5 models, some penalty is incurred by choosing parameters that are believed to better represent the physics upon which the method is founded. The authors feel that such choices allow the SM5CR models to be more reliably applied to solutes outside the training set and to potentials of mean force for the interaction of two solutes than if parameters are chosen solely to minimize the errors over the training set. The SM5CR models give better performance than previous SM*x* models for solutes containing phosphorus because the number of data points in the training set used to fit the phosphorus surface tension coefficients is increased by over 50% by the addition of experimental water–solvent free energies of transfer. The other functional groups that are particularly improved by the new data are amides, ureas, and sulfur compounds.

Acknowledgments

The authors are pleased to acknowledge Tianhai Zhu and David J. Giesen for helpful contributions to this research.

References

1. Giesen, D. J.; Gu, M. Z.; Cramer, C. J.; Truhlar, D. G. J Org Chem 1996, 61, 8720.
2. Hawkins, G. D.; Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. J Phys Chem 1996, 100, 16385.
3. Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. J Org Chem 1998, 63, 4305.
4. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. J Phys Chem B 1998, 102, 3257.
5. Giesen, D. J.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. Theor Chem Acc 1997, 98, 85; 1999, 101, 309(E).
6. Zhu, T.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. J Chem Phys 1998, 109, 9117; 1999, 111, 5624(E).
7. Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Chem Phys Lett 1998, 288, 293.
8. Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. Chem Phys Lett 1998, 288, 293.
9. (a) Zhu, T.; Li, J.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. J Chem Phys 1999, 110, 5503; (b) Chuang, Y.-Y.; Radhakrishnan, M. L.; Fast, P. L.; Cramer, C. J.; Truhlar, D. G. J Phys Chem A 1999, 103, 4893.

10. (a) Tomasi, J.; Persico, M. Chem Rev 1994, 94, 2027; (b) Cramer, C. J.; Truhlar, D. G. Chem Rev 1999, 99, 2161.
11. Cramer, C. J.; Truhlar, D. G. Rev Comput Chem 1995, 6, 1.
12. Rivail, J.-L.; Rinaldi, D. In Computational Chemistry: Reviews of Current Trends; Leszczynski, J., Ed.; World Scientific: Singapore, 1996, Vol. 1, p. 139.
13. Hoijtink, G. J.; de Boer, E.; van der Meij, P. H.; Weijland, W. P. Recl Trav Chim Pays-Bas 1956, 75, 487.
14. Kozaki, T.; Morihashi, M.; Kikuchi, O. J Am Chem Soc 1989, 111, 1547.
15. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J Am Chem Soc 1990, 112, 6127.
16. Cramer, C. J.; Truhlar, D. G. J Am Chem Soc 1991, 113, 8305 [Erratum: 1991, 113, 9901].
17. Klamt, A.; Schüürmann, G. J Chem Soc Perkin Trans 2 1993, 799.
18. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. J Am Chem Soc 1985, 107, 3902.
19. Dewar, M. J. S.; Zoebisch, E. G. J Mol Struct (Theochem) 1988, 180, 1.
20. Dewar, M. J. S.; Yate-Ching, Y. Inorg Chem 1990, 29, 3881.
21. Stewart, J. J. P. J Comput Chem 1989, 10, 221.
22. Thiel, W.; Voityuk, A. A. Int J Quantum Chem 1992, 44, 807.
23. Thiel, W.; Voityuk, A. A. Theor Chim Acta 1992, 81, 391; 1996, 93, 315(E).
24. Thiel, W.; Voityuk, A. A. J Phys Chem 1996, 100, 616.
25. Giesen, D. J.; Storer, J. W.; Cramer, C. J.; Truhlar, D. G. J. J Am Chem Soc 1995, 117, 1057.
26. Bondi, A. J Phys Chem 1964, 68, 441.
27. Lide, D. R., Ed. CRC Handbook of Chemistry and Physics, 75th ed.; CRC Press: Boca Raton, FL, 1995,
28. Abraham, M. H. Chem Soc Rev 1993, 22, 73.
29. Abraham, M. H. J Phys Org Chem 1993, 6, 660.
30. Abraham, M. H.; Claudia, H. S.; Whiting, G. S.; Mitchell, R. C. J Pharm Sci 1994, 83, 1085.
31. Stewart, J. J. P. MOPAC 93.00; Fujitsu Limited: Tokyo, 1993.
32. AMPAC 5.4; Semichem: Shawnee, KS, 1994.
33. Silla, E.; Tuñón, I.; Pascual-Ahuir, J. L. J Comput Chem 1991, 12, 1077.
34. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. J Phys Chem 1996, 100, 19824.
35. Winget, P.; Silverstein, K.; Cramer, C. J.; Truhlar, D. G. unpublished work.
36. Leo, A. J. Medchem Software; BioByte Corp.: Claremont, CA, 1994.
37. Bona, M.; Boudeville, P.; Zerki, O.; Christen, M. O.; Burgot, J. L. J Pharm Sci 1995, 84, 1107.
38. Eliel, E. L.; Wilen, S. H. In Stereochemistry of Organic Compounds; Wiley: New York, 1994.
39. Paglieri, L.; Corongiu, G.; Estrin, D. A. Int J Quantum Chem 1995, 56, 615.
40. Fabian, W. M. F. J Comput Chem 1991, 12, 17.
41. Leszczyński, J. Chem Phys Lett 1990, 174, 347.
42. (a) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. J Comput-Aid Mol Design 1995, 9, 87; (b) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. J Phys Chem A 1998, 102, 1820; (c) Li, J.; Williams, B.; Cramer, C. J.; Truhlar, D. G. J Chem Phys 1999, 110, 724; 1999, 111, 5624(E).
43. The three parameterizations of SM5CR presented in this article were coded in Semichem's AMPAC and submitted to Andrew J. Holder at Semichem, Inc. for inclusion in future releases of Semichem's AMPAC.